
Introduction to the Second Edition

Jan Sobocan

Distracted by the pursuit of wealth, we increasingly ask our schools to turn out useful profit makers, rather than thoughtful citizens

– Nussbaum, 2010, pp. 141-142

Standardized testing is a controversial subject for many reasons. The move to accountability-through-testing gained momentum from the mid 90's through to 2019. In 2002, Ian Wright predicted that more testing would occur, but added that it will be "...the kind of testing that breeds competition rather than measure components of critical thinking" (Wright, p. 149). According to his account, the number of students that would need to be tested in wide scale testing would make it "extremely unlikely that items testing for genuine critical thinking will be administered... Competition seems to be on the upsurge, sometimes in the form of tables, there is a great motivation for teachers to teach to the test".

The many problems that Wright highlights in his 2002 review of the literature on critical thinking and testing persist. In 2011, The Ontario Teachers' Association document "A New Vision for Large-Scale Testing in Ontario," raised many of the same concerns.¹ It concluded that we need to rethink the model of standardized learning first presented in 1994 by the Royal Commission on Learning (RCL) — wherein The Ontario Ministry of Education proposed a standardized model of learning and assessment and implemented the EQAO (Education Quality and Assurance Organization). The move to change the RCL's large-scale testing model vision was based on the argument that standardized testing does not improve teaching, and only evaluates student learning when learning is not occurring in sufficient depth.

When I first organized an international conference to create the first edition of this book, the EQAO and standardized testing was in full force. I saw a need to evaluate such testing because of the negative effects it had on student learning and curricula, encouraging teachers to “teach to the test.” The New Vision document recommends a move from accountability testing because “many features of the EQAO testing [were] redundant and regressive” (p. 2). The careless ways in which “high stakes” tests have been administered and graded, and test results often been employed, have been historically roundly criticized, both in Canada (see Moll’s *Passing the Test: The False Promises of Standardized Testing* [2004]) and in the United States (see Popham’s *The Truth About Testing* [2001]).

Today, Alberta leads all Canadian provinces in the frequency and intensity of government testing, but the value of such testing has been questioned. In 2019, Global News reported that the Alberta Teachers Association president asserted that “PAT [Provincial Achievement Test] results don’t measure creativity — they don’t measure a student’s ability to collaborate, they don’t measure critical thinking. [And] They’re a snapshot of one moment in time” (quote from Alberta Teachers’ Association President Jason Schilling, Heather Yourex-West [2019]). Such limitations raise the question whether the results of such tests can be effectively used to evaluate teachers and programs — particularly in subjects like Language Arts where creativity and critical thinking are said to be primary goals.

The cost of high-stakes testing is another issue as the numbers of classrooms involved in testing regimes rise. In an enterprise as crucial, diverse, and expensive as education, shouldn’t public money be spent in classrooms rather than government testing?

Standardized tests (data gathering instruments commonly referred to as achievement indicators, surveys, and assessments) remain a preferred way to evaluate student learning, curriculum or teacher effectiveness, though the need to incorporate critical thinking into curricula continues to grow. How else can teachers, schools, courses, colleges, programs, provincial, state, and even

national systems of education combat the decline in functional democracy criticized by thinkers like Chomsky [2000].

In our current climate the drive for K-12 high stakes testing and college admissions tests, may be waning. For practical reasons such testing was frequently suspended during the COVID pandemic, and more foundational issues seem to have gained a foothold in public discussion. In 2000, the American K-12 “No Child Left Behind Act” (2000) designed tests similar to Canadian achievement tests, purportedly to “close” achievement gaps. According to many, this policy attempt failed because “...they had long pointed to extensive research showing standardized test scores are most strongly correlated to a student’s life circumstances” (Strauss [2020]). In 2021, a number of states applied to the U.S. Department of Education for waivers that would allow them to forgo the tests.

The problems with high stakes standardized tests do not undermine the suggestion that education should be scrutinized and that teachers and administrators held accountable for the effectiveness of their teaching and programs. More importantly, students struggling in the system should be identified and offered the proper educational support long before they take mandatory high-stakes secondary school exams or university entrance exams such as the SAT. Tests for university admissions, where the financial and emotional stakes for students and their families are very high, necessitate evaluation and better measurement tools. Common sense dictates that increasingly scarce resources should be devoted to teaching that works, and to programs that help struggling learners achieve more than they would if they were not identified.

Standardized tests were and remain attractive because they are the most convenient way to measure the quality of public education. Many administrators and governments support the standardization of learning and assessment, thinking that they ensure teaching and learning quality, at the same time that they identify educational gaps or at-risk students. In these ways, good or valid testing may help sort through difficult questions about teaching and learning. But the value of large-scale testing has

been overestimated, particularly when the ultimate goal is the development of critical or higher-order thinking.

We need to ask how and whether educators can improve large-scale tests to include critical thinking. In the context of changing views and less emphasis on of standardized testing, what alternatives are there to validly test higher-order thinking? I think it is important to begin to answer questions such as these by analyzing the mistakes made in past achievement tests and the use of the data that they generated.

Standardized tests aim, across educational contexts and over time, to assess student knowledge and understanding of a subject in a consistent way. This is a lofty goal, but one that is difficult to achieve with tests that are, more often than not, criticized as the crudest of assessment instruments. In Language Arts — where critical thinking is central in relation to interpretation of texts, Alberta's Diploma exams are dated, and do not include any questions that derive from the novels, poems and short stories which are actually used in classrooms.

One problem is that many of the standardized tests which are used in elementary and post-secondary education are not returned to their takers, or are returned with a reported level only, with no opportunity for students to analyze and learn from their mistakes. When used, such test results assume a distorted importance and are seen as instruments for ranking rather than improvement; causing undue stress; are redundant and regressive; compromise good pedagogy (fostering an educational model driven by teaching to the test); and, cost exorbitant amounts of money which could be redirected into classrooms to reduce class size.

Despite frequent criticism of their cost, the cost of K-12 Achievement tests in Alberta have tripled from \$4 million to \$12 million since the mid-80's. If the tests do not actually promote the current or future learning of higher order thinking skills, what is the worth of the data they produce? One problem is their reliance on multiple choice questions and exercises that measure a very limited range of the content students are learning in their classrooms. The extent to which they measure critical thinking is an issue at a time when the Government of Canada and other

national and international agencies have clearly identified critical thinking as an essential workplace skill. The proper response to this recognition is more research into the ways that K-12 and University educators can incorporate types of questions that measure their idea of critical thinking as “the ability to engage in the process of evaluating ideas or information to reach a rational decision”²

It does not need to be said that certified teachers everywhere are trained at all levels to develop formative tests that accurately measure basic or minimum competencies in their subject areas. This raises the further question whether and how assessments in K-12 education could measure thinking skills? Multiple choice questions have been widely criticized for not soliciting answers or conclusions that can concretely help administrators or teachers address the intricate problems associated with the individual mind or, even more so, the goals of educational policy and practice. Today, it is especially important to foster and validly measure critical thinking in order to address complex issues of social polarization and extremism, and promote international calls for critical thinking programs that might remedy issues related to the decline of democracy.

In many cases, standardized tests have been thought as attractive for precisely the wrong reason: because they can be used to reduce inherently complex questions and information to simplistic arithmetical comparisons — comparisons which are used to rank students, both nationally and internationally. What is required more than ever as the propagandic arm of media replaces argumentation proper, is (at a minimum) tests that better measure cognitive creativity or flexibility. This is an urgent need at a time when thinking has become less complex, and a lack of evidence for claims is nearly absent in the media: “In an increasingly polarizing society, the notion of progress can sometimes feel impossible. Misinformation and the uncompromising way we hold on to our radically different beliefs has divided us... Simply put, we have just stopped thinking” (Fancy [2022]).

Key assumptions about standardized tests and the data they produce ignore key aspects of thinking that should be promoted

in teaching and learning today. Formal and informal evaluations need more depth and adjustment that includes the measurement of media information literacy, creativity, and self-evaluation. In this volume, there are as many test validity questions as there are answers. The second edition of this book aims to illuminate past issues, at the same time that it encourages more thought and research about ethical issues arising from past misinterpretation, misrepresentation, or misuse of data.

Today, the negative consequences of an accountability-through-testing mindset are increasingly apparent. Most importantly, such tests are criticized, not only because they negatively impact students' learning behaviours (promoting memorization rather than thinking), but because they adversely affect the mental health of our youth adversely (Simpson [2016]), and especially at-risk students, and ultimately play too great a role determining the careers of our students. In these ways, the issues discussed in the first edition are still relevant.

The major problem identified in more recent research is the failure of standardized testing when it aims to promote or improve student learning, especially with respect to higher order critical thinking. Further, and equally problematic, is the historical and current suggestion that such testing compromises teaching and the autonomy/ professionalism of teachers, making it more difficult for teachers to focus on broader and more important, but less testable, goals like intellectual development.

Wide-scale testing has not been as zealously pursued in Canada as in the United States, but an increased emphasis on it is one trend in Canadian public education over the last two decades. It has raised the same concerns voiced in response to the American experience. As in the United States, the tests of the past twenty years have been criticized for their design because they were used to collect data in order to catalogue, classify and rank students and schools. Despite these criticisms, test results and rankings remain a matter of intense interest to governments and the general public.

It is ironic that billions of dollars have been spent on standardized testing at a time when "critical thinking" is touted as the fundamental goal of education. The model of education

proposed by those involved in the critical thinking movement suggests that education should aim to endow all citizens with the higher-order thinking skills that will make them critical, self-reflective, and creative participants in democracy. So conceived, education should NOT endeavour to produce students who possess specific circumscribed knowledge and information just so they can be tested more easily, in the interests of “accountability.”

Instead, the goal should be students who have more life-relevant (though much more difficult to define) thinking abilities, skills, and dispositions — i.e., students who are disposed to ask questions, to reason through issues and problems, and to self-evaluate. Such students will be more able to acquire and assess new knowledge and information, but it is difficult to see how their abilities will be abilities that can be tested with instruments as undeveloped as those containing only multiple-choice and short-answer items accompanied by rigid, inside-the-box (correct or incorrect) scoring criteria.

The chapters of this book examine topics at the intersection of an emerging commitment to the idea that critical thinking should be the central goal of education and international debates about testing and educational accountability. They consider, among many others, the following topics and issues:

- different accounts of critical thinking and different approaches to its testing and assessment;
- the effects of testing on students;
- the criteria for judging the validity of test instruments and testing contexts;
- the validity (or invalidity) of particular, widely used performance or standardized achievement tests that claim, in part or whole, to measure critical thinking;
- the policy issues around the testing of higher-order thinking; and
- the relationship between critical and creative thinking, and how we might assess creativity.

As the authors of these chapters demonstrate, a commitment to critical thinking as a central goal of education intensifies the issues raised by standardized testing and assessment.

A critical scrutiny of historical attempts to teach and assess critical thinking is especially important when one considers the limited progress made in testing design over the past 17 years. This editor's view is the same as it was in the first edition of this book — that higher-order thinking skills (in particular, informal logic or everyday reasoning skills) need to be better taught in schools and tested with a greater degree of validity. It is less clear how these instructional and curricular goals can be achieved. In contexts plagued by competition, ranking, and grade inflation, the essential question is whether and how we can know what works if we do not have some ways to measure our successes and, equally importantly, our failures?

The issues raised here are not limited to K-12 education. Most colleges and universities declare a strong commitment to critical thinking and its development in their courses and programs, but few turn this rhetoric into a concerted attempt to ensure that critical thinking shapes their curriculum. More frequently, those supporting traditional programs protect themselves from change by claiming that their program already embodies the spirit and goal of critical thinking. Such claims are ironic, not just because they are made without any evidence of an understanding of the critical thinking literature, but because they are made without any serious attempt to marshal evidence in their favour, i.e., in a manner that fundamentally violates the central components of critical thinking.

Some progress has been made in the critical thinking courses that are a staple in undergraduate education in the arts. Though I view this as a positive development, others have questioned about the efficacy of such courses. Theirs is a legitimate concern because it cannot be said that the efficacy of the courses has been proven or backed by extensive research and testing. To make matters worse, philosophy courses and philosophy departments that emphasize the reasoning skills courses in their curricula have waned considerably.

The content of the reasoning skills courses taught in universities varies widely, reflecting fundamental disagreements about the best way to teach critical thinking (from the point of view of logic, dialectics or rhetoric, or via some mix of their approaches). Instead of consensus among the experts, one finds conflicting approaches that represent the particular biases of the individual instructors — some still emphasizing traditional formal logic, some focusing on fallacies, some employing rhetorical techniques, and so on.

From this point of view, it is somewhat paradoxical that the claimed value of university courses in critical thinking has been backed by vague truisms and prejudices in favour of the value of critical thinking, and not by critical reflection that genuinely demonstrates either the relevance of such courses for learners, or the effectiveness of teaching.

From a broader point of view, the claims that universities make about their commitment to critical thinking are sometimes suspect. One of Canada's best liberal arts institutions publishes a recruitment "viewbook," a website and a calendar that repeatedly touts the ability to think critically as one of the benefits of its degrees. In its calendar, for example, one reads that its programs shape "leaders who are critical thinkers, problem solvers and creative participants in society."

These are laudable ideals but it is difficult to see how they have, in any conscious way, shaped the programs in question. There is no explicit program of the sort that Don Hatcher describes in his contribution to this book (i.e., the Baker University liberal arts program, which has critical thinking as an explicit and detailed goal). And nothing that someone who has studied critical thinking (which has been an area of research and scholarship for over thirty years) would recognize as a concerted effort to infuse critical thinking into the curriculum. Rather, the university (in a manner inconsistent with the critical reflection that is the heart of critical thinking) operates with the expectation that its programs fulfill this ideal. This is not because the university is more negligent than other universities in this regard, but because a rhetorical, not a substantive, commitment to critical thinking is the norm in most liberal arts programs in North American universities.

This makes it all the more important that we re-examine the success of different attempts to assess higher-level thinking. In the course of that examination, and especially in a context that concerns passing and failing grades, and the attempt to assess the human mind for its strengths and weaknesses, it is important to consider questions raised by both theorists of critical thinking and experts in assessment. What parts of a critical thinking process need to be tested to establish that a person is thinking at a higher level? How can teaching and assessment tools incorporate critical thinking? How can assessment be done in a way that stretches assessment beyond students' basic abilities? What are the best kinds of assessment tools for doing so? What role, if any, can standardized testing play? And, is formal testing even necessary in an attempt to decide how teachers, disciplines, schools, colleges, universities, and whole systems of education can best embrace critical thinking as a true goal rather than a mere platitude?

The issues raised in this book reflect the complexity of the issues raised by the testing of critical thinking. They include difficulties establishing the nature and definition of critical thinking, the ethics of assessment policy and practice, and the impact of assessment on how we teach critical thinking. Such issues are so complicated that many commentators (including several authors in this volume) believe that the skills and/ or dispositions that make up critical thinking are in principle too complex to be captured and quantified in a standardized testing format. Whether one goes this far or not, it is difficult to find ways to validly test critical thinking and especially difficult to test systems of education, as standardized formats require simplified scoring keys and come at great cost to taxpayers.

Many of the authors who have contributed to this book have responded to the need to find ways to more accurately test critical thinking skills and dispositions — skills and dispositions that are, I believe, essential for the maintenance of democracy. Considerations of this sort give rise to many questions. In many ways, the one that best captures the issues discussed in this book is the question how negative conclusions about testing programs

might be reconciled with the recognition that concerns about educational accountability are legitimate.

Many more specific questions of validity are spawned by this general query.

- How can we establish whether students are acquiring the traits that characterize the critical thinker?
- How can we establish the extent to which critical thinking is taught in the K-12 and post-secondary curricula?
- How can we ensure some consistency between instruction/student learning and critical thinking as an educational goal?
- How can we use multiple-choice instruments more effectively to measure critical thinking?
- What is to be said about past and existing tests?
- What happens to teaching when tests become the measure of successful teaching in the classroom?
- What other kinds of classroom assessments and evaluations can be used to measure critical thinking, and are they better measures of it?
- On what basis should we choose between different approaches to critical thinking programs and courses?
- How can we ensure that the comparative gains data we use to inform curriculum development are reliable?
- How can we ensure that the data we collect is used democratically, to improve student learning?

The contributors to this volume have addressed these and related questions from a variety of perspectives. Different authors have focused on different components of education as professors, educational theorists, philosophers, evaluation experts, and policy and program developers. Many of them have taught and/or administered (and continue to teach and/ or administer) critical thinking instruction in K-12, or at the university level.

The first edition of this book aimed to situate it within the extensive research literature that has spurred the development of critical thinking and cognate disciplines (informal logic, argumentation theory, rhetoric, dialectics, etc.). In the last twenty years, they have made the study of such thinking a promising intellectual exercise focused on the ways in which we think and reason, the ways in which we should think and reason, and the ways in which we can best teach students to be stronger thinkers. This new edition aims to continue a discussion of educational theory and practice that will better integrate the study and teaching of critical thinking. Hopefully this will motivate educators, governments, and theorists to work together to redress some of the historical disappointments that have attended earlier efforts, and that still exist today.

One notable obstacle to progress in this area has been a continuing debate over the definition of “critical thinking” — a debate engaged by the authors of many of the essays to follow. In some ways, issues of definition, and the debates they produce, enrich our understanding of the nature and teaching of critical thinking. Discussions of the definition of critical thinking have, for example, contributed to a growing recognition that it should be expanded to incorporate literacy in general, and media literacy in particular. The recognition of the latter importantly includes a critique of technology, with an emphasis on the images and Internet advertisements that bombard us every day. This broader understanding of the content of critical thinking can usefully promote a still more significant mandate for critical thinking education, one that many of the contributors here have passionately pursued.

We should not expect a consensus on some exact definition of critical thinking (least of all from those who work as philosophers, who are prone to disagree about definitions). Complete agreement is not a prerequisite for a better understanding and assessment of critical thinking and its pedagogy. However one defines critical thinking, everyone agrees that it encompasses certain core abilities and practices — the ability to evaluate a range of views and evidence, to recognize and deal fairly with opposing points of

view, to ask key questions, and to self-reflect. Understanding critical thinking in these general terms provides what is needed when we try to study it from both theoretical and pedagogical points of view.

The chapters included in this book have been organized in Parts that represent key issues and themes that arise when one considers critical thinking and its assessment. In Part One the validity of various popular standardized tests is examined. In Part Two, the authors discuss often overlooked issues with respect to the relationship between critical thinking and creative thinking (because critical thinking, in the proper sense of the term, implies something more than the ability to be critical of others' points of view). In Part Three, particular approaches to critical thinking teaching and assessment are discussed. Here, the authors discuss different programs and related evaluations of their success to teacher education, classroom instruction, and to non-standardized informal assessments of critical thinking. Part Four includes attempts to answer broad questions about critical thinking education policy or accountability, and the ways in which how such policy supports (or does not support) critical thinking education and testing. In the final Part of the book, Sharon Murphy comments on all of the essays in the book, suggesting a way to further our thinking on issues of critical thinking and assessment.

I envision this book as a volume that does something more than criticize (though criticism is an essential and a healthy component of critical thinking). As educators, we want to move beyond negative criticism toward critical decision-making. I have tried to develop the 2nd edition in a way that will allow its readers — philosophers, administrators, educational theorists, teachers, students, policy-makers, and others — to emerge with a better understanding of critical thinking and its relationship to historical issues of testing and assessment.

I hope for something more: that this book becomes an important historical look at standardized testing and presents a shared understanding of how critical thinking might be better taught and tested. In the future, this might include the design of better and more reliable tests that could have an impact on curriculum and

policy to an extent that motivates us all to promote critical thinking education. If our tests were valid measures of critical thinking, then teaching to these tests would be a good idea.

References

- Apple, M. 2001. *Educating the “right” way: Markets, standards, God and inequality*. New York: Routledge Falmer.
- Bullen, M. 1998. Participation and critical thinking in online university distance education. *Journal of Distance Education* 13(2): 1-32.
- Chomsky, N. 2000. Democracy and education. In *Chomsky on miseducation*, ed. D. Macedo, 37-55. Oxford: Rowman & Littlefield.
- Darling-Hammond, L. 1999. *Teacher quality and student achievement: A review of state policy evidence*. Seattle: Center for the Study of Teaching and Policy.
- Fancy, N. 2022. Every high school student should have to take a philosophy course. *CBC News Opinion*, Feb 04, <https://www.cbc.ca/news/opinion/opinion-high-school-philosophy-1.6331790>.
- Gorrie, P. 2004. Literacy test a write-off? *The Toronto Star*, Sunday, February 15.
- Kuhn, D. 1991. *The skills of argument*. Cambridge: Cambridge University Press.
- Lehrer, J. 2004. *A newshour with Jim Lehrer* transcript. Online focus. Newsmaker: Hans Blix March 17, 2004. http://www.pbs.org/newshour/bb/international/jan-june04/blix_3-17.html.
- Moll, M. 2004. *Passing the test: The false promises of standardized testing*. Ottawa: Canadian Centre for Policy Alternatives.
- Nixon, G. 1999. Whatever happened to ‘heightened consciousness’? *Journal of Curriculum Studies* 31(6): 625-33.
- Owens, A. 2002. Putting schools to the test: How standardized exams are changing education in Canada. *National Post*, Saturday, November 16.
- Popham, W. 2001. *The truth about testing*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Runte, R. 1998. The impact of centralized examinations on teacher professionalism. *Canadian Journal of Education* 23: 166-81.
- Ricci, C. 2004. Breaking the silence: An EQAO marker speaks out against standardized testing. *Our Schools, Our Selves Winter*: 75-88.
- Simpson, C. 2016. *Effects of Standardized Testing on Students' Well-Being*. Harvard Graduate School of Education.
https://projects.iq.harvard.edu/files/eap/files/c._simpson_effects_of_testing_on_well_being_5_16.pdf
- Strauss, V. 2020. It looks like the beginning of the end of American's obsession with standardized tests. *The Washington Post*, June 21.
- van Gelder, T. 2005. Teaching critical thinking: Some lessons from cognitive science. *College Teaching* 53(1): 41-6.
- Wright, I. 2002. Critical Thinking in the Schools: Why Doesn't Much Happen? A review of the Literature. *Informal Logic* 22(2): pp 137-154).
- Yourex-West, H. 2019. Why standardized tests are a controversial subject for Alberta Schools. *Global News*, posted Sept 3, 2019.

Notes

1. https://www.otffeo.on.ca/en/wp-content/uploads/sites/2/2013/09/new_vision.pdf
2. from the Skills Canada website: <https://skillscanadanl.ca/programs-events/essential-skills/#:~:text=What%20are%20the%209%20Essential,Thinking%2C%20Document%20Use%2C%20Digital> (last accessed April 9/2022)